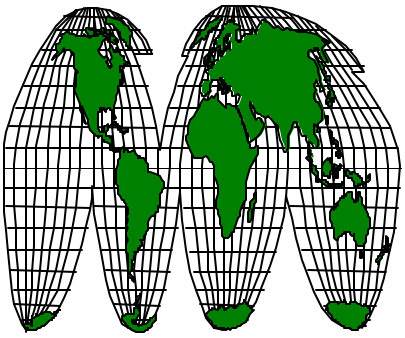
Data Grid Projects



Harvey B Newman
California Institute of Technology

Grid Meeting
CERN, September 27, 2000

Roles of Projects for HENP Distributed Analysis

- RD45, GIOD — Networked Object Databases
- Clipper/GC FNAL/SAM — High speed access to Objects or File data for processing and analysis
- SLAC/OOFS — Distributed File System + Objectivity Interface
- NILE, Condor: — Fault Tolerant Distributed Computing
- MONARC — LHC Computing Models: Architecture, Simulation, Strategy, Politics
- ***PPDG*** — ***First Distributed Data Services and Data Grid System Prototypes***
- ALDAP — OO Database Structures & Access Methods for Astrophysics and HENP Data
- ***GriPhyN*** — ***Production-Scale Data Grids***
- ***EU Data Grid***

SDSS Data Grid (In GriPhyN)

Three main functions:

Raw data processing on a Grid (FNAL)

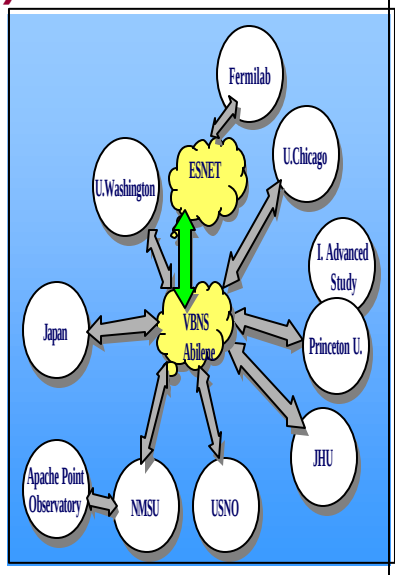
- ➔ Rapid turnaround with TBs of data
- ➔ Accessible storage of all image data

Fast science analysis environment (JHU)

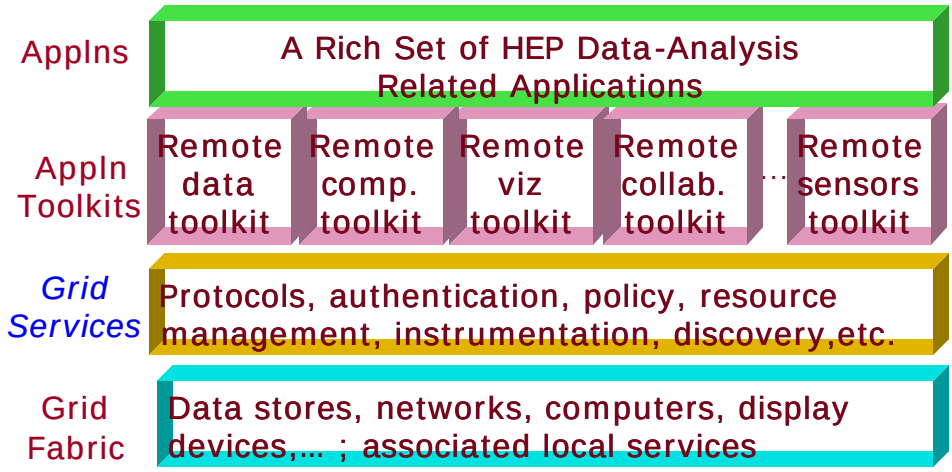
- ➔ Combined data access + analysis of calibrated data
- ➔ Distributed I/O layer and processing layer; shared by whole collaboration

Public data access

- ➔ SDSS data browsing for astronomers, and students
- ➔ Complex query engine for the public



Grid Services Architecture [*]



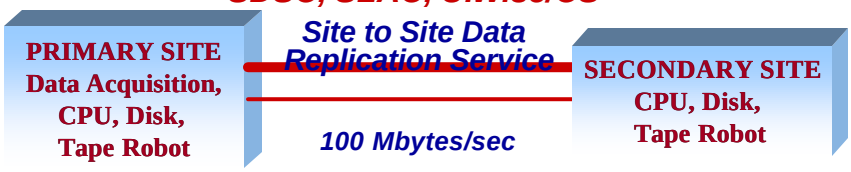
[*] Adapted from Ian Foster: there are computing grids, access (collaborative) grids, data grids, ...

Data Grids: Better Global Resource Use and Faster Turnaround

- **Build Information and Security (Authentication + Authorization) Infrastructures**
- **Coordinated use of computing, data handling and network resources through:**
  - ➔ ***Data caching, query estimation, co-scheduling, transaction management***
  - ➔ ***Network and site "instrumentation": performance tracking, monitoring, problem trapping and handling***
  - ➔ ***Robust Transactions (Agents)***
    - & ***Redirection; error recovery; fallback***

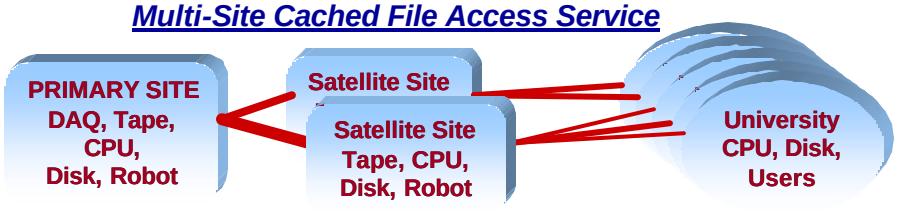
The Particle Physics Data Grid (PPDG)

ANL, BNL, Caltech, FNAL, JLAB, LBNL, SDSC, SLAC, U.Wisc/CS



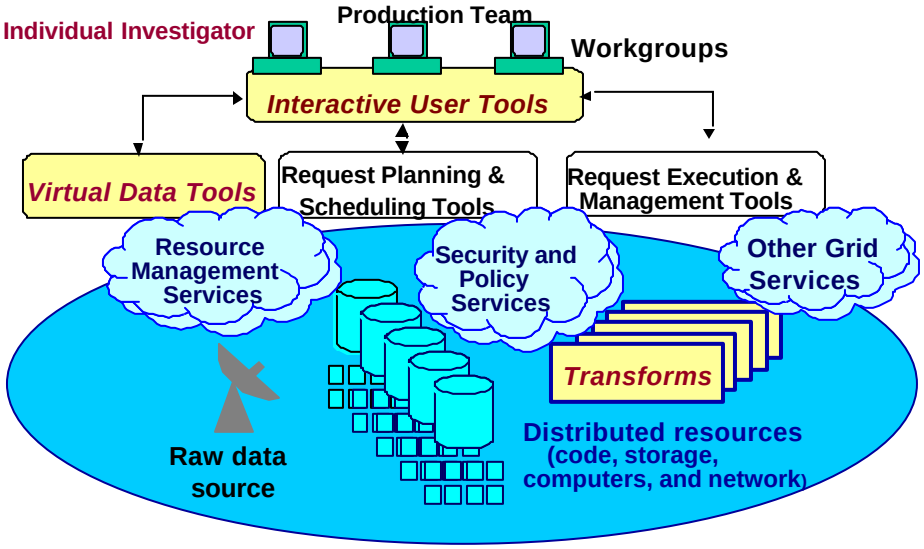
- First Round Goal: Optimized cached read access to 10-100 Gbytes drawn from a total data set of 0.1 to ~1 Petabyte
- Matchmaking, Co-Scheduling: SRB, Condor, Globus services; HRM, NWS

Multi-Site Cached File Access Service



GriPhyN: PetaScale Virtual Data Grids

Build the Foundation for Petascale Virtual Data Grids



EU-Grid Project Work Packages

Work Package Number	Work Package title	Lead contractor
WP1	Grid Workload Management	INFN
WP2	Grid Data Management	CERN
WP3	Grid Monitoring Services	PPARC
WP4	Fabric Management	CERN
WP5	Mass Storage Management	PPARC
WP6	Integration Testbed	CNRS
WP7	Network Services	CNRS
WP8	High Energy Physics Applications	CERN
WP9	Earth Observation Science Applications	ESA
WP10	Biology Science Applications	INFN
WP11	Dissemination and Exploitation	INFN
WP12	Project Management	CERN

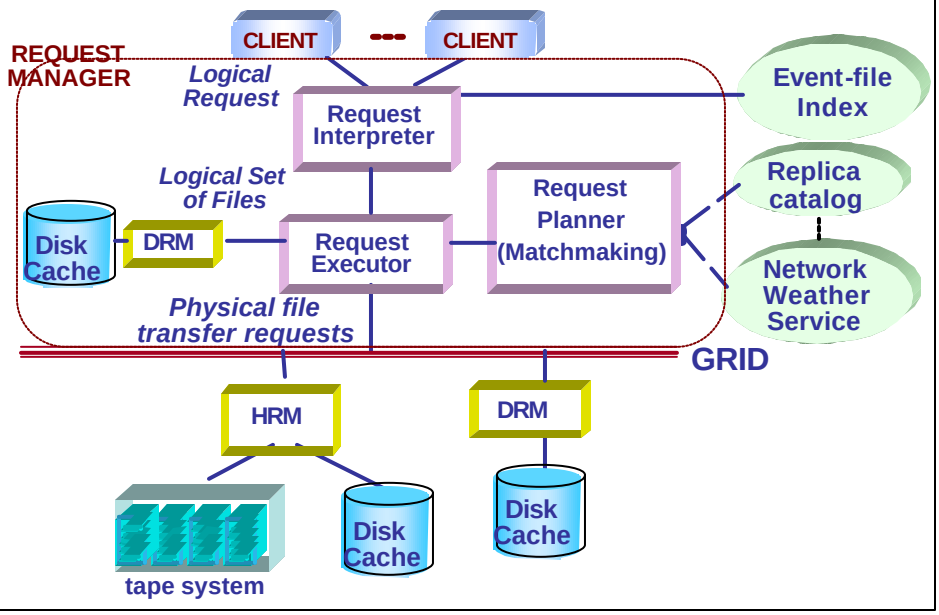
Grid Project Status

- **PPDG: Happroved for one more year**
- **GriPhyN: $ 12M IT R&D Approved for 5 Years; $ 58 M to Go: Hardware, Ops Support, Networking**
- **EU DataGrid: People for R&D Only**

PPDG Work at Caltech (1)

- **Data Grid Development**
  - **High throughput data transfer (JB, HN, AS)**
  - **Globus Security and Information Infrastructure (AS, MH)**
  - **Distributed Data Management (JB, HN, AS, MH, KH)**
    - ***GRID DATA MANAGEMENT PROTOTYPE (GDMP) V1.0, With EU DataGrid WP2, CMS/CERN, FNAL***
  - **Distributed Computing & Task Scheduling (KH, TH, VL, AS, MH)**
  - **Tier2 Center design (HN, JB; with UCSD)**
  - **Data Structures and Clustering (KH, JB, HN)**
  - **Distributed System Simulations (IL; KH, HN)**

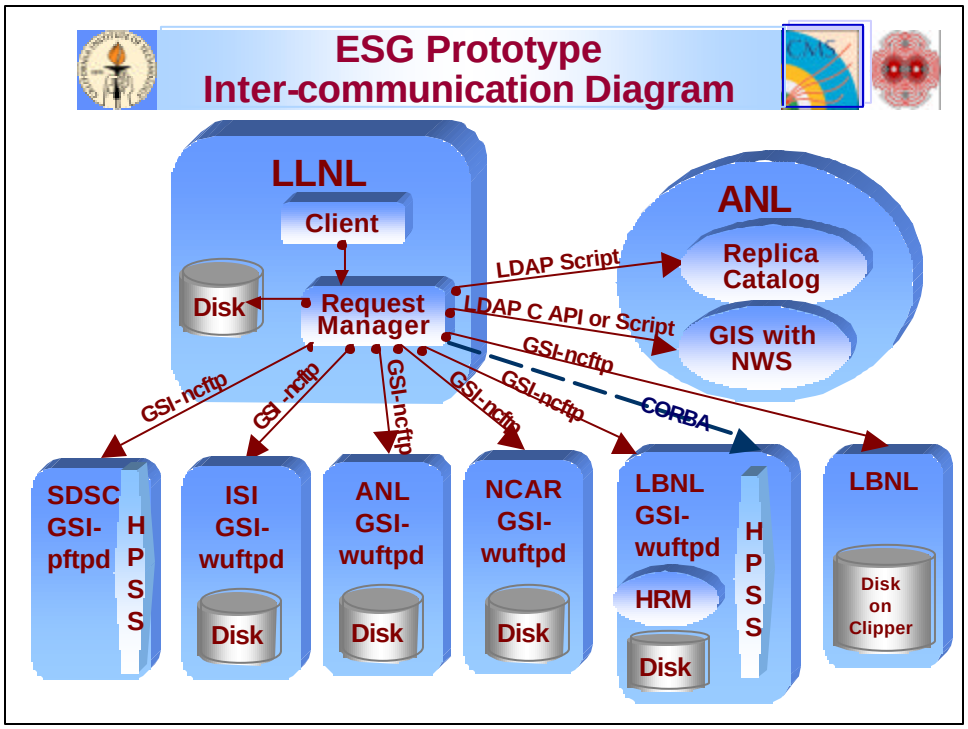
PPDG WG1: Request Manager

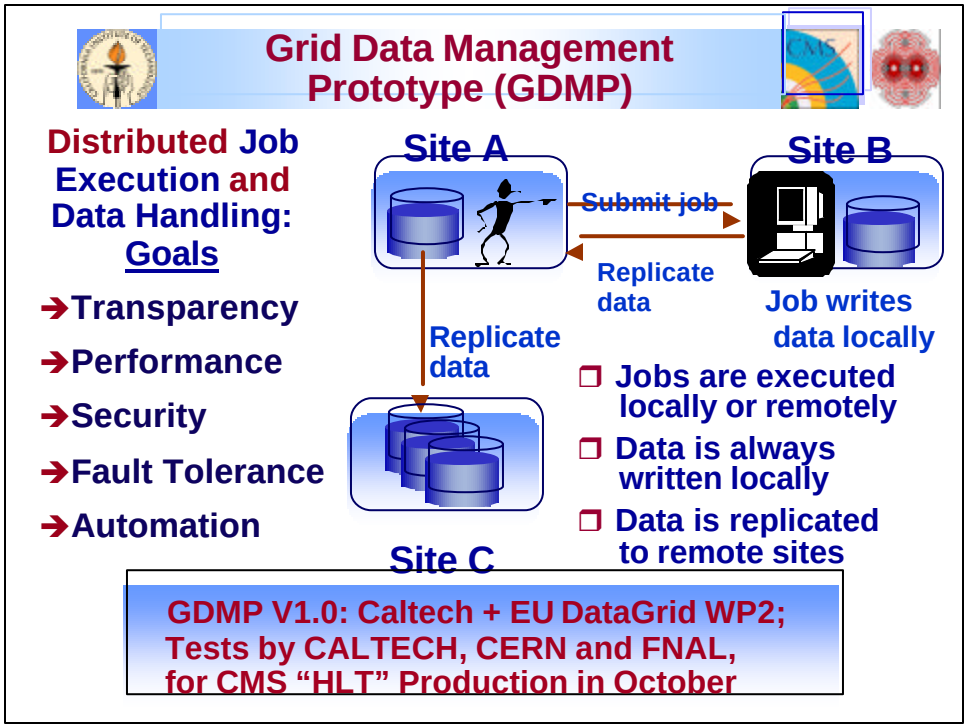


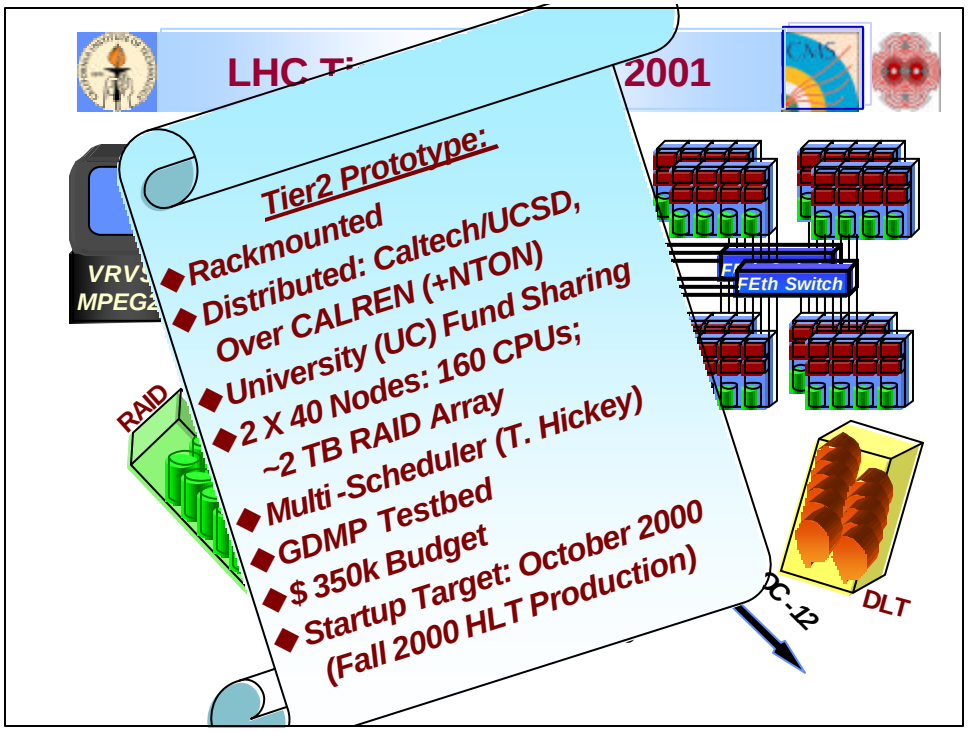
Earth Sciences Grid Prototype: LBNL, Uwisc, SDSC, ANL,...

Request Manager (ReqM) is newly developed software at LBNL

- ➔ accepts a request to cache a set of logical file names
- ➔ checks for each replica location
- ➔ gets for each replica location NWS bandwidth
- ➔ selects “lowest” cost location
- ➔ initiates transfer using GSI-FTP
- ➔ monitors progress, responds to status command


ESG Prototype
Inter-communication Diagram
LLNL
Client
Disk
Request Manager
ANL
Replica Catalog
GIS with NWS
LDAP Script
LDAP C API or Script
GSI-ncftp
CORBA
SDSC GSI-pftpd
HPSS
ISI GSI-wuftpd
Disk
ANL GSI-wuftpd
Disk
NCAR GSI-wuftpd
Disk
LBNL GSI-wuftpd
HRM
Disk
HPSS
LBNL
Disk on Clipper


Grid Data Management
Prototype (GDMP)
Distributed Job Execution and Data Handling:
Goals
➔Transparency
➔Performance
➔Security
➔Fault Tolerance
➔Automation
Site A
Site B
Submit job
Replicate data
Job writes data locally
Replicate data
Site C
❒ Jobs are executed locally or remotely
❒ Data is always written locally
❒ Data is replicated to remote sites
GDMP V1.0: Caltech + EU DataGrid WP2;
Tests by CALTECH, CERN and FNAL,
for CMS "HLT" Production in October



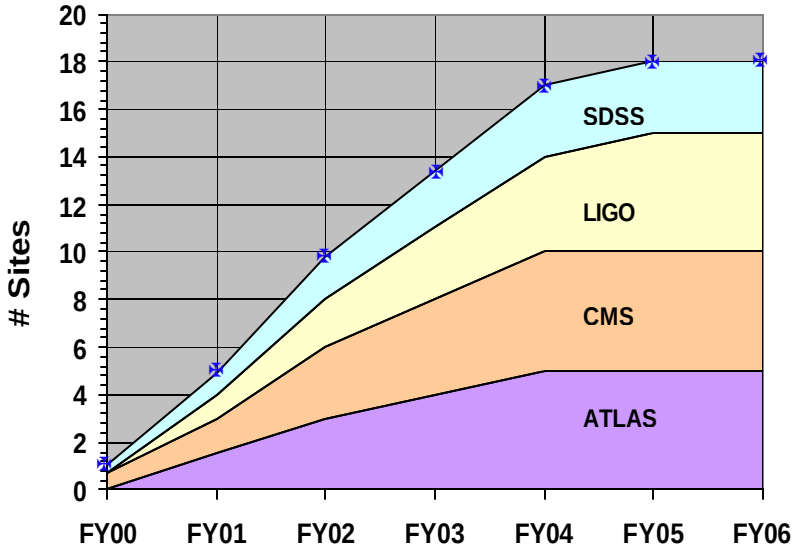
LHC Tier2 Architecture and Cost

➔ Linux Farm of 128 Nodes (256 CPUs + disk)	$ 350 K
➔ Data Server with RAID Array	$ 150 K
➔ Tape Library	$ 50 K
➔ Tape Media and Consumables	$ 40 K
➔ LAN Switches	$ 60 K
➔ Collaborative Tools & Infrastructure	$ 50 K
➔ Installation & Infrastructure	$ 50 K
➔ Net Connect to WAN (Abilene)	$ 300 K
➔ Staff (Ops and System Support)	$ 200 K ✦
➔ Total Estimated Cost (First Year)	$1,250 K
➔ Average Yearly Cost including evolution, upgrade and operations ✠	$ 750K

✦ 1.5 – 2 FTE support required per Tier2

✠ Assumes 3 year hardware replacement

Tier2 Site-Units vs Time



Common Grid Data Management Issues

- Data movement and responsibility for updating the Replica Catalog
- Metadata update and replica consistency
  - Concurrency and locking
- Performance characteristics of replicas
- Advance Reservation: Policy, time-limit
  - How to advertise policy and resource availability
- Pull versus push (strategy; security)
- Fault tolerance; recovery procedures
- Queue management
- Access control, both global and local

PPDG Relationship to GriPhyN

The PIs of PPDG (Mount, Newman) and of GriPhyN (Avery, Foster) will set-up a Coordination Board to ensure:

- That PPDG Facilities and the results of PPDG experience are available to GriPhyN.
- As new tools are created within GriPhyN, they will be evaluated by PPDG.

Specific developments planned in FY 2000 - 2001

- Development of a generalized file-mover framework.
- Implementation/generalization of a metadata catalog, resource broker, resource managers.
- Implementation of transparent write access for files.
- Implementation of limited support for "agents".
- Implementation of distributed resource management for the Data Grid.
- instrumentation of all Data Grid components in support of a systematic approach to measurement of and modeling of Data Grid behavior

Grid Project Convergence

- Ongoing Joint Work with WP2 (AS, KH)
- Work Starting with WP5 (SE)
- Exchanges: HS and KS contacts in US
- PPDG/Globus Meetings + Discussion
- Meeting of PPDG, DataGrid and GriPhyN managements at ACAT2000 (FNAL)
- Joint meeting of the Projects planned for early December (December 8 ?)